Claims

1. A file server system including

a first file server including a file server change memory;

a second file server including a file server change memory;

a mass storage element;

said first file server and said second file server being coupled to said mass storage element;

means for copying a descriptor of a file system change to both said first and second file servers, whereby said first file server processes said file system change while said second file server maintains its copy of said descriptor in its file server change memory; and

means for said second file server to perform a file system change in its file server change memory in response to a service interruption by said first file server.

2. A system as in claim 1, including at least one said mass storage element for each said file server.

3. A system as in claim 1, wherein a first said file server is disposed for processing said file system changes atomically, whereby a second said file server can on failover process exactly those file system changes not already processed by said first file server.

4. A system as in claim 1, wherein a first said file server is disposed to respond identically to service interruptions for itself and for a second said file server.

5. A system as in claim 1, wherein at least one said file server is disposed to delay output to said mass storage element without delaying a response to file system changes.

6. A system as in claim 1, wherein at least one said file server responds to a file system change before committing a result of said file system change to mass storage.

7. A system as in claim 1, wherein

each one of said file servers is coupled to at least a portion of said file server change memory using local memory access; and

each one of said file servers is coupled to at least a portion of said file server change memory using remote memory access

8. A system as in claim 1, wherein said descriptor includes a file server request.

9. A system as in claim 1, wherein said file server change memory includes a disk block.

10. A system as in claim 1, wherein said file server change memory includes a file server request.

11. A system as in claim 1, wherein said file server change memory is disposed to delay output to said mass storage element without delaying a response to file server requests.

12. A system as in claim 1, wherein

said mass storage element includes a file storage system;

each said file server is disposed for leaving said file storage system in an internally consistent state after processing file system changes;

said internally consistent state is associated with a set of completed file system changes;

said set of completed file system changes is identifiable by each said file server.

13. A system as in claim 1, wherein said mass storage element includes a file storage system and each said file server is disposed for leaving said file storage system in an internally consistent state after processing each said file system change.

14. A file server system as in claim 1, wherein

said mass storage element includes a primary mass storage element and a mirror mass storage element; and

said first file server processes said file system change for both said primary mass storage element and said mirror mass storage element.

15. A system as in claim 1, wherein said means for copying includes access to at least one of said first and second file server change memories using a NUMA network.

16. A system as in claim 1, wherein said means for copying includes remote memory access to at least one of said first and second file server change memories.

17. A system as in claim 1, wherein said means for said second file server to perform a file server request in its file server change memory is also operative in response to a service interruption by said second file server.

18. A file server system including

a first file server coupled to a first set of mass storage devices;

a second file server coupled to a second set of mass storage devices;

a server change memory;

said first file server disposed for receiving a file server request and in response thereto copying a descriptor of a file system change into said server change memory; and

said first file server disposed for processing said file system change for both said first set of mass storage devices and for at least one said mass storage device in said second set.

19. A system as in claim 18, wherein

said second file server is disposed for receiving a file server request and in response thereto copying a descriptor of a file system change into said server change memory; and

said second file server is disposed for processing said file system change for both said second set of mass storage devices and for at least one said mass storage device in said first set.

20. A system as in claim 18, wherein said server change memory includes a disk block.

21. A system as in claim 18, wherein said server change memory includes a file server request.

22. A system as in claim 18, wherein said server change memory includes a first portion disposed at said first file server and a second portion disposed at said second file server.

23. A system as in claim 18, wherein

said server change memory includes a first portion disposed at said first file server and a second portion disposed at said second file server; and

said first file server is disposed for copying said descriptor into both said first portion and said second portion.

24. A system as in claim 18, wherein

said server change memory includes a first portion disposed at said first file server and a second portion disposed at said second file server; and

said first file server and said second file server are each disposed for copying said descriptor into both said first portion and said second portion.

25. A system as in claim 18, wherein said server change memory is disposed to delay output to said mass storage element without delaying a response to file server requests.

26. A file server system including

a plurality of file servers, said plurality of file servers coupled to a mass storage element and at least one file server change memory;

each said file server disposed for receiving a file server request and in response thereto copying a descriptor of a file system change into said file server change memory; and

each said file server disposed for responding to a service interruption by performing a file system change in said file server change memory.

27. A system as in claim 26, including at least one said mass storage element for each said file server.

28. A system as in claim 26, including at least one said server change memory for each said file server.

29. A system as in claim 26, wherein a first said file server is disposed for processing said file system changes atomically, whereby a second said file server can on failover process exactly those file system changes not already processed by said first file server.

30. A system as in claim 26, wherein a first said file server is disposed to respond identically to service interruptions for itself and for a second said file server.

31. A system as in claim 26, wherein at least one said file server delays output to said mass storage element without delaying a response to file server requests.

32. A system as in claim 26, wherein at least one said file server responds to a file system change before committing a result of said file system change to mass storage.

33. A system as in claim 26, wherein

each one of said file servers is coupled to at least a portion of said file server change memory using local memory access; and

each one of said file servers is coupled to at least a portion of said file server change memory using remote memory access.

34. A system as in claim 26, wherein each said file server is disposed for copying said descriptors using a NUMA network.

35. A system as in claim 26, wherein each said file server is disposed for copying said descriptors using remote memory access.

36. A system as in claim 26, wherein said file server change memory includes a disk block.

37. A system as in claim 26, wherein said file server change memory includes a file server request.

38. A system as in claim 26, wherein said file server change memory is disposed to delay output to said mass storage element without delaying a response to file server requests.

39. A system as in claim 26, wherein said mass storage element includes a file storage system and each said file server is disposed for leaving said file storage system in an internally consistent state after processing each said file system change.

40. A system as in claim 26, wherein

said mass storage element includes a file storage system;

each said file server is disposed for leaving said file storage system in an internally consistent state after processing file system changes;

said internally consistent state is associated with a set of completed file system changes;

said set of completed file system changes is identifiable by each said file server.

41. A file server system as in claim 26, wherein

said mass storage element includes a primary mass storage element and a mirror mass storage element; and

said first file server processes said file system change for both said primary mass storage element and said mirror mass storage element.

42. A method of operating a file server system, said method including steps for

responding to an incoming file server request by copying a descriptor of a file system change to both a first file server and a second file server;

processing said file system change at said first file server while maintaining said descriptor copy at said second file server; and

performing, at said second file server, a file system change in response to a copied descriptor and a service interruption by said first file server.

43. A method as in claim 42, including steps for associating a first file server and a second file server with a mass storage element.

44. A method as in claim 42, including steps for delaying output by at least one said file server to said mass storage system without delaying a response to file system changes.

45. A method as in claim 42, wherein a first said file server is disposed for processing said file system changes atomically, whereby a second said file server can on failover process exactly those file system changes not already processed by said first file server.

46. A method as in claim 42, wherein a first said file server is disposed to respond identically to service interruptions for itself and for a second said file server.

47. A method as in claim 42, wherein at least one said file server responds to a file system change before committing a result of said file system change to mass storage.

48. A method as in claim 42, wherein

each said file server includes a file server change memory;

each one of said file servers is coupled to at least a portion of said file server change memory using local memory access; and

each one of said file servers is coupled to at least a portion of said file server request memory using remote memory access.

49. A method as in claim 42, wherein said file server change memory includes a disk block.

50. A method as in claim 42, wherein said file server change memory includes a file server request.

52. A method as in claim 42, wherein said file server change memory is disposed to delay output to said mass storage element without delaying a response to file server requests.

53. A method as in claim 42, wherein said mass storage element includes a file storage system and each said file server is disposed for leaving said file storage system in an internally consistent state after processing each said file system change.

54. A method as in claim 42, wherein said steps for performing a file system change in response to a copied descriptor are also operative in response to a service interruption by said second file server.

55. A method as in claim 42, wherein said steps for processing includes steps for processing said file system change at both a primary mass storage element and a mirror mass storage element.

56. A method of operating a file server system, said method including steps for

receiving a file server request at one of a plurality of file servers and in response thereto copying a descriptor of a file system change into a server change memory;

processing said file system change for both a first set of mass storage devices coupled to a first one said file server and for at least one said mass storage device in a second set of mass storage devices coupled to a second one said file server.

56 57. A method as in claim 56, wherein said descriptor includes a file server request.

57 58. A method as in claim 56, wherein said server change memory includes a disk block.

58 59. A method as in claim 56, wherein said server change memory includes a file server request.

59 60. A method as in claim 56, wherein said server change memory includes a first portion disposed at said first file server and a second portion disposed at said second file server.

60 61. A method as in claim 56, wherein said server change memory includes a first portion disposed at said first file server and a second portion disposed at

said second file server; and wherein said steps for copying include steps for copying said descriptor into both said first portion and said second portion.

62. A method as in claim 56, wherein said server change memory includes a first portion disposed at said first file server and a second portion disposed at said second file server; and said steps for copying include steps for copying said descriptor into both said first portion and said second portion by either of said first file server or said second file server.

63. A method as in claim 56, wherein said server change memory is disposed to delay output to said mass storage element without delaying a response to file server requests.

64. A method as in claim 56, wherein

said steps for receiving include receiving a file server request at either said first file server or said second file server, and said steps for copying said descriptor include copying by either said first file server or said second file server; and including steps for

processing said file system change for both said second set of mass storage devices and for at least one said mass storage device in said first set.

64 ~~65~~. A method of operating a file server system, said method including steps for

receiving a file server request at one of a plurality of file servers and in response thereto copying a descriptor of a file system change into a file server change memory; and

responding to a service interruption by performing a file system change in response to a descriptor in said file server change memory.

65 ~~66~~. A method as in claim 65, including steps for associating a plurality of file servers with at least one mass storage element and at least one file server change memory.

66 ~~67~~. A method as in claim 65, including steps for delaying output to a mass storage element without delaying a response to file server requests.

67 ~~68~~. A method as in claim 65, including steps for leaving a file storage system on said mass storage element in an internally consistent state after processing each said file system change.

68 ~~69~~. A method as in claim 65, including steps for

leaving a file storage system on said mass storage element in an internally consistent state after processing file system changes;

associating said internally consistent state with a set of completed file system changes; and

identifying said set of completed file system changes by at least one said file server.

69

~~70.~~ A method as in claim 65, including steps for performing said received file server request at both a primary mass storage element and a mirror mass storage element.

10 ~~71.~~ A method as in claim 65, including steps for

processing said file system changes atomically at a first said file server; and

on failover processing exactly those file system changes not already processed by said first file server.

11 ~~72.~~ A method as in claim 65, including steps for responding identically at a first said file server to service interruptions for itself and for a second said file server.

12 ~~73.~~ A method as in claim 65, wherein said file server change memory includes a disk block.

13 ~~74.~~ A method as in claim 65, wherein said file server change memory includes a file server request.

~~75~~ 74. A method as in claim 65, wherein said file server change memory is disposed to delay output to said mass storage element without delaying a response to file server requests.

~~76~~ 75. A method as in claim 65, including steps for responding to a file system change before committing a result of said file system change to mass storage at one said file server.